Proc. Indian natn. Sci. Acad., **54**, A, No. 3, 1988 pp. 354–364.

STATISTICAL MODELS OF CLIMATE RECONSTRUCTION USING TREE RING DATA

G B PANT, K RUPA KUMAR *and* H P BORGAONKAR

Indian Institute of Tropical Meteorology, Shivajinagar, Pune 411 005, India

(*Received 3 February 1988*)

After obtaining precisely dated and replicated chronology of standardised ring-width indices, the variations in ring-width data are related to variations in climatic data. This process, known as calibration, uses mathematical or statistical procedures to transfer growth measurements into climatic estimates. Multivariate techniques with different levels of complexity are presently in use in dendroclimatic studies. Principal components are used to overcome the problem of interrelationships among the independent variables, by converting the raw data into uncorrelated eigenvectors. Stepwise multiple regression procedures are employed to select the eigenvectors significantly contributing to the variance in the ring-width data, and thereby obtaining the response functions indicating the effect of climatic variables at different stages of the growing season. Stepwise multiple regression techniques are also used in selecting the predictands to be used to develop the calibration equations. The results of the calibration are then statistically verified on independent data before attempting the climate reconstruction.

This paper discusses these techniques and the computational procedures involved in reconstruction of past climates from tree rings. An example of reconstructing the All India Summer Monsoon rainfall over the past 4 centuries is presented.

Key Words : Dendroclimatology; Climate Reconstruction; Statistical Models

INTRODUCTION

VARIATIONS in the widths of the annual rings of trees have long been recognised as an important proxy source of past climatic information. Considerable work has been done, extensively in North America and Europe in building up precisely dated chronologies of tree ring widths as well as densities for various species responsive to climatic variations. These chronologies have been related to climate during the period of instrumental records and then used in relationships in the past also. The process of estimating the past climate, involving calibration, verification and reconstruction, requires the use of empirical/statistical techniques, mainly because of the lack of quantitative knowledge on the physiological response of trees to climate. Thus, the data on climate and tree rings are allowed to speak for themselves and the useful signals are picked up.

Several statistical techniques have been adopted to extract the maximum possible information from the tree rings. These methods are under continuous refinement, incorporating the various developments in computational procedures.

This paper presents an overview of the basic approach involved and the statistical techniques employed in modelling tree-ring and climate relations. An example of reconstructing the all India Summer Monsoon rainfall over the past 4 centuries from an indirect source of dendroclimatic data, is also presented.

The dendroclimatology group at the Indian Institute of Tropical Meteorology is actively pursuing the collection of tree ring samples from various parts of the country and attempts are being made to develop suitably cross-matched chronologies of ring widths. Several samples have already been collected from the Kashmir Valley, Thane and some parts of Andhra Pradesh. Preliminary analyses have shown good promise of the possibility of reconstructing the past climate of Indian region.[1–5] In this connection, statistical response function analyses and stepwise multiple regression modelling have been under consideration to be applied to the data. The basic structure of the model was developed by Fritts[6] and several additions and modifications have been made by the dendroclimatology group at the Lamont Doherty Geological Observatory at the University of Columbia, USA.

RING WIDTH DATA

The frequency domain of the tree ring chronology can be visualised as composed of four different kinds of signals[7] :

$$R(t) = C + B + D + E, \quad \ldots(1)$$

where $R(t)$ is measured ring width in year t, C is the macroclimatic signal common to trees at a site, B is the biological growth curve as a function of increasing age, D is the tree disturbance signal due to factors such as competition effects, forest fires and pests, and E is the random growth signal unique to each specimen. In order to maximise the climate signal, it is necessary to recognise and remove or control the other parameters.

Standardisation is the process by which the non-climatic signals of biological growth trend and/or tree disturbance are removed from the ring width data. This process has the inherent assumption that the slowly varying components of the ring width variation are mainly due to non-climatic factors. Standardisation is done by fitting an appropriate curve to the ring width series and then calculating a new time series:

$$\text{Index}\,(t) = R\,(t)/y\,(t), \quad \ldots(2)$$

where $y(t)$ is the expected yearly growth determined from curve fitting. The curves usually fitted are negative exponential or a polynomial of suitable degree. Negative exponential is very efficient in removing the juvenile growth effects. The cubic spline has been found to give better results than the polynomial, particularly for samples taken from forest interiors, where the trees environment is subject to considerable changes due to competition effects etc.[8] In the process of standardisation, certain amount of climatic information is likely to be lost where the signal and noise spectra overlap in the lower frequencies. This problem can be minimised by comparing the fitted curves of different tree ring series; any consistent low frequency similarities between the curves would suggest that a common signal, perhaps climate, has been removed. However, general disturbances such as fire or insect infestation affecting large stands of trees complicate this approach.

The random noise in the ring width series is minimised by averaging cross matched chronologies of several samples of the same species from a single site. However, at sites where the tree ring samples display heterogeneous variations due to various factors, the average series may lose considerable climatic information. In such cases, principal components analysis is performed over all the chronologies to divide them into several homogeneous subsets on the basis of the amplitudes of the first eigenvector.[9] Such techniques help in developing chronologies that correlate better with local climatic data than the site-averaged chronology.

Climatic Data

Statistical models of reconstructing past climates are used to quantify climatic relation with the ring width variation during the period of instrumental record. The models use variables such as monthly or seasonal precipitation or temperature, the total annual water balance, the number of hours that temperatures are above a critical value, the number of days in the growing season, or some other expression of the growth controlling conditions that can be derived from the available data. The climatic data at an observatory reasonably close to the site of tree ring sampling is generally used for the analysis. If, however, sufficient network of observatories is present in the neighbourhood of the site, regionally averaged climatic data are likely to provide more satisfactory results.[10]

Calibration, Verification and Reconstruction

After obtaining a precisely dated and replicated chronology of standardised ring width indices, their relationships with the climatic data are studied using a suitable mathematical or statistical procedure, to identify the climatic parameters that can be reconstructed from ring width data. This process of developing a scheme to convert growth measurements into terms of climatic estimates is known as *calibration.*

At the primary level of calibration is the simple linear regression model with only two variables: growth indices and a climatic parameter. However, this model requires an over-simplified assumption that the climatic variable selected is the only one accounting for most of the variance in the tree growth record. As the tree growth is a more complex process encompassing several climatic interactions, a more objective approach is the use of multiple regression techniques. Such techniques help in selecting from a variety of climatic variables those which are primarily responsible for the variance in the tree growth record. The analysis results in an equation expressing the response of the tree to variations in the most important climatic variables, and this is known as a *response function.* In practice, response functions are always multivariate, reflecting the complexity of the tree growth relationship.

One of the difficulties in multiple regression is the fact that the climatic variables are themselves often highly correlated. A way round this problem is to express the variance of climatic data in terms of principal components or eigenvectors and to use these as predictors in the regression procedure. Principal components analysis involves statistical transformation of the original (interrelated) data set to produce a set of orthogonal (uncorrelated) eigenvectors. The primary eigenvectors can be thought of as preferred modes of distribution of the data set and account for most of its variance. The value or amplitude of each eigenvector will vary from year to year being highest in the year when that particular combination of climatic variables which the eigenvector represents, is most apparent. Conversely

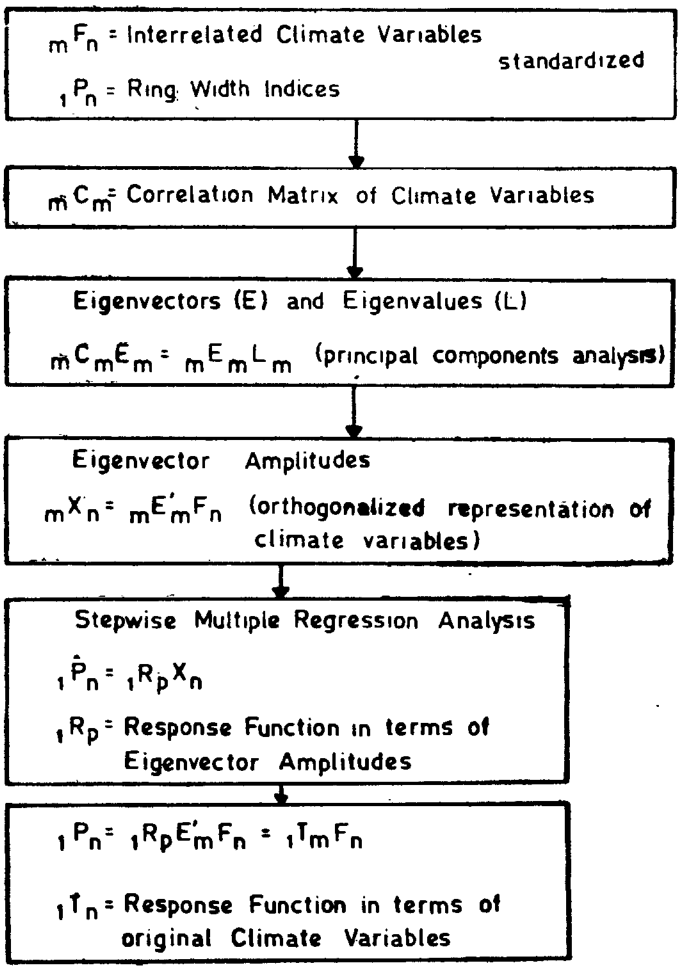

FIG 1 Schematic illustration of the procedure for obtaining response function.

it will be lowest in the year when the inverse of this combination is most apparent. By using eigenvector amplitudes as independent variables in the stepwise regression procedure, a higher proportion of the dependent data variance can be accounted for by few variables than would be possible using the original climatic data set. Once the regression coefficients have been calculated, the eigenvectors incorporated in the regression equation are mathematically transformed into a new set of coefficients corresponding to the original set of variables. These coefficients are termed as weights or elements of the response function, representing the response of the tree to the combination of climatic conditions represented in the eigenvector.[6] The process of obtaining response function as described above is schematically presented in Fig. 1.

The response functions focus on the relationship between tree growth on an individual site and its response to the climate in the area. Similar methodology can be applied to study the way in which a network of trees responds to a specific climatic parameter. In this case, variance of the tree growth data is expressed in the form of eigenvectors, each one thus representing a spatial pattern of growth variations. Amplitudes of these eigenvectors are then used as independent variables in the multiple regression analysis. The resulting equation is termed as *transfer function* whereby spatial patterns of growth records are transferred into climatic estimates.

There are several approaches in calculating the response functions with slight modifications depending upon the requirements. Berger *et al.*[11] proposed the removal of persistence from the whole chronology before obtaining the response functions. Introducing the climate of prior growth seasons is another possibility which can give better results if there are any lag effects on the tree. Guiot[12] proposed a somewhat different approach called spectral multivariate analysis which involves studying the response functions separately under different frequencies of the spectrum, using digital filters (high-pass and low-pass filters, complementary to each other). This method makes an important distinction between influences of short term and long term climatic variations on the tree growth.

The relationship between the dendroclimatic indictor x (predictor set) and the climate variable, y (predict and set) may be expressed as

$$y = f(x) + e,$$

where f is a linear function and e, an error component which includes non-modelled physiological and climatic interactions. Multiple linear regression methods with least squares criteria are used to estimate the function $f(x)$. The regression estimates are then tested to determine their statistical significance. To select the predictor variables to be included in the regression function, there are several schemes, which use different criteria based on the statistical properties of the relationships.[13] Stepwise multiple regression analysis selects the variables one by one, with the condition that the added variable must give the highest improvement in multiple correlation with least effect on the reduction of statistical significance due to reduced degrees of freedom.

The verification, which must be independent in nature to be more reliable, may be done by comparison with other proxy sources of past climatic information or with independent predictand observations. The latter is a more objective approach. In this approach, a part of the tree ring-climate data set is kept aside for testing the calibrated function. Since these data are not used to develop the calibrated function, it is expected to provide an independent assessment of the equation regarding its usefulness in reconstruction.

After selecting a calibrated equation and after a suitable independent verification, the equation is then used to reconstruct the past climate. A series is thus prepared showing the variation of the climatic variable as reconstructed from the tree ring data. However, before making any interpretations of the variations in the series, serious attention should be paid to the statistical nature of the calibration and verification results. It must be kept in mind that the reconstructed series is only that part of the climate as explained by tree rings and has its own limitations. Any conclusion made from the series must give due concern to such limitations.

Reconstruction of All India Summer Monsoon Rainfall Over the Last 400 Years

Some of the above statistical modelling techniques have been applied to reconstruct the All India Summer Monsoon rainfall from the year 1602, using an indirect dendroclimatic data set. This is presented here for demonstration purposes only and discussion of palaeoclimatic implications of the results are out of the scope of this paper. The data used, techniques employed and the results are briefly described below.

(*i*) *Data Base*

Mooley and Parthasarathy[14] have prepared a long homogeneous series of All India Summer Monsoon (June to September) rainfall, which is an area weighted average of the rainfall at 306 well distributed plain stations. The data for the period 1871–1959 have been used for the present study. Lough and Fritts[15] have reconstructed Wright's[16] index of the Southern Oscillation (SO) back to 1600 A D using tree ring chronologies from both western North America and the Southern Hemisphere. Variations of the Southern Oscillation which comprises a large scale fluctuation of atmospheric mass between the southeastern Pacific and western Pacific/Indian Ocean regions, with a see-saw surface pressure oscillation between the two areas, is known to influence surface climate over a large part of the globe. The influence of SO on the Indian Summer Monsoon has been extensively studied and the results conclusively establish a significant relation between the two, though it does not provide any predictive value for monsoon.[17–20] The reconstructed series are available for four seasons, namely, December to February (DJF), March to May (MAM), June to August (JJA), and September to November (SON), during the period 1602 to 1960.[21] In the present study, these series are used as an indirect source of dendroclimatic data related to the Indian Summer Monsoon and a reconstruction is attempted using the methods described above.

(ii) Response Function

In order to derive the response function of monsoon rainfall to the SO indices of the four seasons, it is necessary to consider lead and lag effects also, keeping in view the nature of teleconnections between the two. Hence, for each seasonal series, three lags are considered, viz. − 1, 0 and + 1 representing preceding, current and following years. Thus, 12 different series are obtained, three each for the four seasons. These 12 series are subjected to principal components analysis to account for the inter-relationships. The data of reconstructed series were taken for the period 1870 to 1960 and because of the lead and lagged series, the effective data length is 89 years, 1871 to 1959.

The results of the principal components analysis are summarised in Table I. It may be seen that the first 4 eigenvectors alone are accounting for 92 per cent of the total variance which is obviously due to the high inter-correlations between the series.

TABLE I

Principal components of reconstructed SO indices of 4 seasons each with —1, 0 and + 1 lags during 1871–1959

Principal Component Number	Eigenvalues	Cumulative variance explained (%)
1	4.0149	33.46
2	3.3482	61.36
3	2.0632	78.55
4	1.7176	92.87
5	0.4713	96.79
6	0.1406	97.97
7	0.1305	99.05
8	0.0846	99.76
9	0.0253	99.97
10	0.0025	99.99
11	0.0011	100.00
12	0.0002	100.00

These 12 eigenvectors are taken as the independent variables and are subjected to a stepwise multiple regression with the monsoon rainfall. The stepping was terminated after entering the first 4 eigenvectors, as the remaining eigenvectors account for a very small amount of variance. The regression equation with 4 eigenvectors (multiple correlation = 0.41) is then transformed into the terms of original variables, giving a response function as shown in Fig. 2. It can be seen that the SO index of JJA and SON of the current year and DJF and MAM of the subsequent year are significantly related to the Indian Summer Monsoon. This agrees well with the findings of different studies on the effects of SO on monsoon.[18,22]

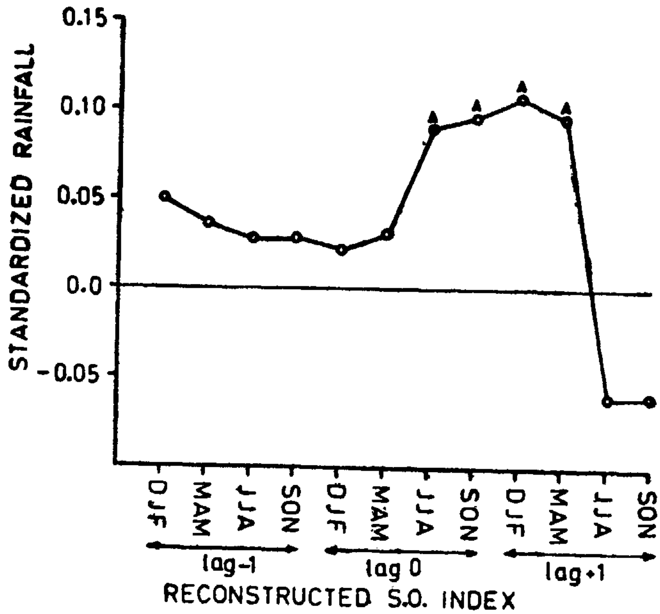

FIG 2 Response function of All India Summer Monsoon rainfall to reconstructed SO index (1871–1959) represented by 4 eigenvectors (R^2 = 0.41). Crowns over the circular points indicate significant responses at 5 per cent level.

The authors have also subjected the original SO index of Wright[16] to the similar response function analysis. In this case also, a high proportion of the variance (87 per cent) is accounted for by the first 4 principal components. The response function of monsoon rainfall to Wright's original SO index (multiple correlation = 0.67) is presented in Fig. 3. This is similar to the response function with the reconstructed index with some differences in the magnitudes of responses. This indicates that the reconstructed index displays similar relationship to the monsoon as the original SO index, thus providing a logical basis for proceeding with the reconstruction of monsoon rainfall.

(*iii*) *Calibration, Verification and Reconstruction*

For calibrating the reconstructed SO index on monsoon rainfall, stepwise multiple regression was employed. For this, data for the period 1871–1939 has been used. The remaining data during 1940–1959 have been kept aside for independent verification of the calibrating equation. From an examination of different calibrating equations under different steps, the equation containing the seasonal indices of SON of preceding year, DJF and JJA of the current year and MAM and SON of the following year emerged as the best equation in terms of calibration as well as verification. Here, it may be noticed that, the index of SON of current year and DJF of the following year did not enter the regression which was mainly due to their high correlation with the indices already entered. Thus, addition of

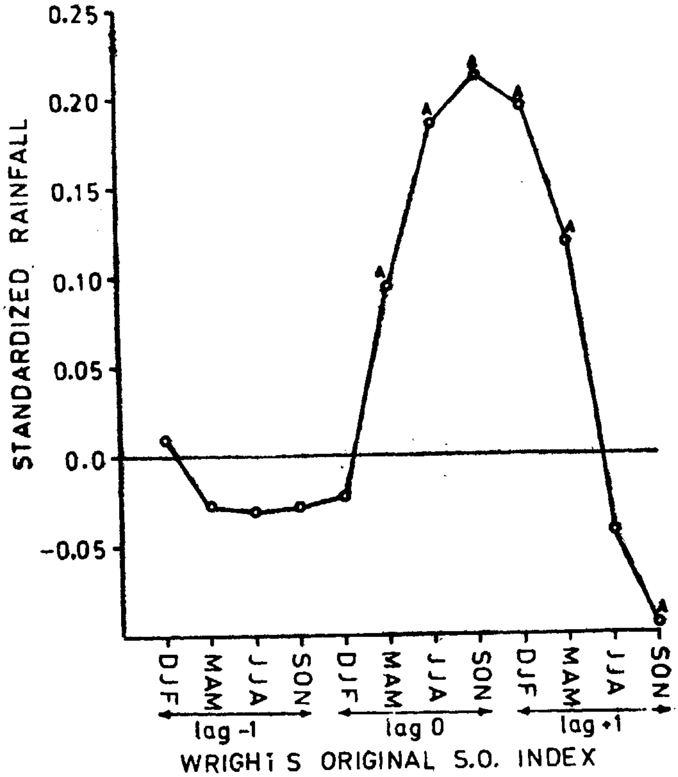

FIG 3 Response function of All India Summer Monsoon rainfall to Wright's original SO index (1871–1959) represented by 4 eigenvectors (R^2 = 0.67). Crowns over circular points indicate significant response at 5 per cent level.

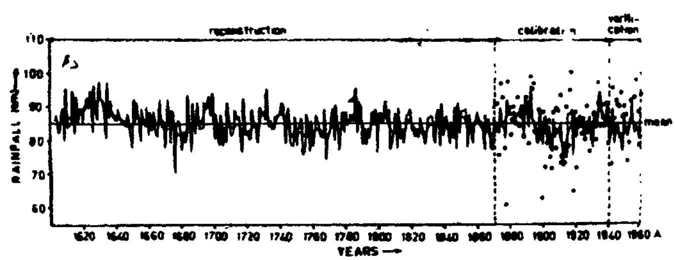

FIG 4 Reconstructed series of All India Summer Monsoon rainfall for the period 1602 to 1960 along with the 9-point Gaussian low pass filter (smooth curve). The circular points indicate actual rainfall.

these variables would not improve the calibration equation. The details of the calibration equation are given below:

Calibration Period—1871–1939

No. of years:	69	
No. of degrees of freedom:	Total :	68
	Regression :	5
	Residual:	63
Multiple correlation coefficient (R) =		0.574
R^2 =		0.329
Adjusted R^2 =		0.276
F–ratio =		6.1905

(Significant at 0.1 per cent level)

Verification period: *1940–1959*

No. of independent data points: 20.

Correlation between Reconstructed and Original Rainfall = 0.3452 (Significant at 20 Per cent Level)

Using this equation, All India Summer Monsoon rainfall has been reconstructed for the period, 1602 to 1870. The rainfall series during the period of reconstruction, calibration and verification along with a 9–point Gaussian low pass filter is presented in Fig. 4. This indicates that there was an excess rainfall epoch during the period 1610 to 1635. Other less prominent epochs are also seen in the series which reflect the basic nature of the series as revealed by the instrumental record during the present century, i. e., the occurrence of different epochs.[14] More frequent droughts during a major part of the 19th century, as shown by the reconstructed series, have also been noted by Mooley and Pant[23] who have compiled the information from historical and other sources. However, since this reconstruction is based on teleconnections alone, corroborative evidence from local data is needed before making any conclusive statement.

Acknowledgements

The authors are grateful to Shri D R Sikka, Director, IITM, for encouragement and interest in this area of research. The authors are indebted to Professor H C Fritts, Tree Ring Research Laboratory, University of Arizona, USA for kindly providing the reconstructed data on the Southern Oscillation index. Thanks are due to the National Informatics Centre (Western Region), Pune, for permission to use their NEC S-1000 computer system.

References

1. G B Pant Role of tree ring analysis and related studies in palaeoclimatology: preliminary survey and scope for Indian region *Mausam* **30** (1979) 439
2. G B Pant Climatological signals from the annual growth rings of selected tree species of India *Mausam* **34** (1983) 251
3. G B Pant Dendroclimatic studies on the northern trees In: *Climate and Geology of Kashmir and Central Asia* (*Eds* D P Agrawal *et al.*) Today and Tomorrow's Printers and Publishers New Delhi (1984) 115

4. G B Pant and H P Borgaonkar Growth rings of teak trees and regional climatology In: *Environmental Management* (*Eds* L R Singh R C Tiwari and R P Srivastava) The Allahabad Geographical Society Allahabad, India (1983) 153
5. G B Pant and H P Borgaonkar Growth rate of Chir Pines (*Pinus roxburghii*) trees in Kumaon area in relation to regional climatology *Himal Res Dev* **3** (1984) 1
6. H C Fritts *Tree Rings and Climate* Academic Press London (1976)
7. D A Graybill Chronology development and analysis In: *Climate from Tree Rings* (*Eds* M K Hughes P M Kelly J R Pitcher V C LaMarche Jr) Cambridge University Press Cambridge (1982) 21
8. E R Cook and K Peters The smoothing spline: a new approach to standardising forest interior tree ring width series for dendroclimatic studies *Tree Ring Bull* **41** (1981) 45
9. K Peters G C Jacoby and E R Cook Principal components analysis of tree ring sites *Tree Ring Bull* **41** (1981) 1
10. T J Blasing D N Duvick and D C West Dendroclimatic calibration and verification using regionally averaged and single station precipitation data *Tree Ring Bull* **41** (1981) 37
11. A L Berger J Guliot L Mathieu and A V Munaut Tree rings and climate in Morocco *Tree Ring Bull* **39** (1979) 61
12. K Guiot *Special Multivariate Regression in Dendroclimatology* Contribution No 21 Institute d' Astronomie et de Geophysique, Universite Catholique de Louvain—la—Neuve (1980)
13. G R Lofgren and J H Hunt Transfer functions In: *Climate from Tree Rings* (*Eds* M K Hughes P M Kelly J R Pitcher V C LaMarche Jr) Cambridge University Press Cambridge (1982) 50
14. D A Mooley and B Parthasarathy Fluctuations in All India Summer Monsoon rainfall during 1871–1978 *Climatic Change* **6** (1984) 287
15. J M Lough and H C Fritts The Southern Oscillation and Tree Rings : 1600-1961 *J Climate Appl Met* **24** (1985) 952
16. P B Wright *An Index of the Southern Oscillation* Rep No CRU-RP4 Climatic Research Unit Univ East Anglia Norwich UK (1975) 22pp
17. G B Pant and B Parthasarathy Some aspects of an association between the Southern Oscillation and Indian Summer Monsoon *Arch Met Geoph Biokl Ser* **B29** (1981) 245
18. B Parthasarathy and G B Pant The spatial and temporal relationships between the Indian Summer monsoon rainfall and the Southern Oscillation *Tellus* **36**–A (1984) 269
19. B Parthasarathy and G B Pant Seasonal relationships between Indian Summer Monsoon rainfall and the Southern Oscillation *J Climatol* **5** (1985) 369
20. D A Mooley B Parthasarathy and N A Sontakke Relationship between All India Summer Monsoon rainfall and Southern Oscillation/eastern equatorial Pacific sea surface temperature *Proc Indian Acad Sci* (*Earth Planet Sci*) **94** (1985) 199
21. J M Lough *The Southern Oscillation and North American Climate, 1602 to 1962 Tech Note* No 24 Laboratory of Tree Ring Research University of Arizona Tucson Arizona USA (1983) 43 pp
22. H N Bhalme and S K Jadhav The Southern Oscillation and its relation to the monsoon rainfall *J Climatol* **4** (1984) 509
23. D A Mooley and G B Pant Droughts in India over the last 200 years, their socio-economic impacts and remedial measures for them In: *Climate and History-Studies in Past Climates and Their Impact on Man* (*Eds* T M L Wigley *et al*) Cambridge University Press Cambridge (1981) 465